

APLICACION DE UN PERCEPTRON MULTICAPA A LA ESTIMACION DE ANGULOS DE LLEGADA DE FUENTES DE BANDA ESTRECHA.

José Ramón Cerquides Bueno, Juan Fernandez Rubio
Dpto Teoría de la señal y comunicaciones. ETSIT-BARCELONA. UPC¹

ABSTRACT

This paper shows how a multilayered perceptron can be applied to DOA estimation, avoiding many of the problems of the classical methods. A review of the main characteristics of the perceptron is made, and several algorithms are used to train the network. At the recall phase, the system is able to respond quite well to situations not present at the training set.

INTRODUCCION

En esta comunicación se describen la estructura y los principios básicos de funcionamiento del perceptrón, se presentan diversas modalidades del algoritmo de aprendizaje, y se estudia la capacidad de aplicación de una red de este estilo a la estimación de la dirección de llegada de una serie de fuentes a un array de sensores.

EL PERCEPTRON MULTICAPA. ESTRUCTURA Y PROPAGACION INPUT-OUTPUT.

Un perceptrón multicapa es un tipo especial de red neuronal. [1] Su capacidad de proceso se basa en una densa interconectividad entre sus nodos procesadores y en un elevado paralelismo presente en todo momento en la estructura.

El perceptrón puede representarse como una sucesión de capas (fig.1). La capa 0 es la de entrada, las capas 1..C-1 son capas ocultas o intermedias, y la capa C es la de salida.

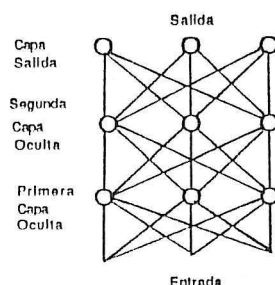


Figura 1. Estructura del perceptron

Cada capa contiene un conjunto de nodos, cada uno de los cuales se caracteriza por un vector:

$$W_l^c ; c=1..C=\text{número de capa} ; l=1..N_c=\text{número de nodo}$$

siendo N_c el número de nodos en la capa c.

Las componentes de los vectores de nodo serán:

$$W_{lj}^c ; j=1..N_{c-1}+1$$

¹Este trabajo ha sido financiado parcialmente por S.A.E.S. y parcialmente por el Plan Nacional de Investigación de las Tecnologías de Información y Comunicaciones.

es decir, hay tantas componentes como nodos tenga la capa anterior, mas una adicional, utilizada para conservar el valor del umbral de excitación. Cada capa, además, tiene asociado un vector:

$$o_i^c ; i=1..N_c$$

La entrada al nodo i de la capa $c+1$ se calcula, entonces, como:

$$\sum_{j=1}^{N_c} \left(W_{ij}^{c+1} o_j^c \right) - W_{iN_{c-1}+1}^{c+1}$$

Puede observarse que resulta más cómodo trabajar añadiendo una componente -1 a los vectores o . De ese modo, la entrada al nodo i será:

$$W_{i1}^{c+1} o_i^c$$

y la salida de ese nodo vendrá dada por:

$$o_i^{c+1} = F \left(W_{i1}^{c+1} o_i^c \right)$$

donde $F(x)$ es una función que relaciona entrada y salida.

Se puede ver que para $c=0$; o^0 es el vector de entrada (o datos de partida), y que o^c son los datos finales suministrados por la red.

EL ALGORITMO DE ENTRENAMIENTO

Antes de que la red sea útil es necesario someterla a un proceso de aprendizaje, a fin de calcular los valores óptimos de los vectores de nodo. Para ello, la técnica a seguir es presentar un elevado número de veces una secuencia de datos a la entrada, para la cual conocemos la salida. A continuación se describe brevemente el proceso.

Si disponemos de un conjunto de muestras de entrada $I^1..I^m$, y conocemos sus correspondientes salidas $O^1..O^m$, tendremos, al propagarse por el interior de la red la entrada k , una salida o^c distinta, en general, de O^k .

En general, podemos definir un error extendido al conjunto de muestras:

$$\epsilon = \sum_{k=1}^m | o_k^c - O^k |^2$$

Ahora, para calcular los pesos óptimos, basta con que nos movamos en dirección opuesta al gradiente de la función de error anteriormente definida.

Así, en la iteración $n+1$ tendremos:

$$W_{ij}^c [n+1] = W_{ij}^c [n] - \alpha \frac{\partial \epsilon}{\partial W_{ij}^c}$$

El cálculo de las derivadas parciales puede hacerse por capas, comenzando por la última y propagando hacia atrás. Por ello se conoce este

algoritmo con el nombre de backpropagation.

Existe la posibilidad de modificar el método, a fin de acelerar la convergencia durante la fase de entrenamiento. Se comentan, a continuación, las dos técnicas más utilizadas para ello:

a) Hacer servir un término de inercia, procedente de las anteriores correcciones, y que va amortiguándose en iteraciones sucesivas. En este caso se obtiene:

$$w_{ij}^c [n+1] = w_{ij}^c [n] - \alpha \frac{\partial \mathcal{E}}{\partial w_{ij}^c} + \beta \left(w_{ij}^c [n] - w_{ij}^c [n-1] \right)$$

b) Intentar que el factor de ajuste se adapte lo mejor posible a las características del gradiente de la función de error. La nueva expresión es [2]:

$$w_{ij}^c [n+1] = w_{ij}^c [n] - \alpha_{ij}^c [n] \frac{\partial \mathcal{E}}{\partial w_{ij}^c}$$

tomando como criterio de modificación de α_{ij}^c el siguiente:

$$\alpha_{ij}^c [n+1] = \begin{cases} s \alpha_{ij}^c [n] & \text{si } \nabla_{ij}^c \mathcal{E} [n] \nabla_{ij}^c \mathcal{E} [n-1] > 0; s > 1; \\ b \alpha_{ij}^c [n] & \text{si } \nabla_{ij}^c \mathcal{E} [n] \nabla_{ij}^c \mathcal{E} [n-1] < 0; b < 1; \end{cases}$$

de modo que dos iteraciones que mantengan el mismo signo en el gradiente producen un incremento del factor α . Si por el contrario, hay un cambio de signo en dicho gradiente, α se reduce, para evitar que el sistema pueda oscilar o hacerse inestable.

ESTIMACION DE ANGULO DE LLEGADA

El problema de la estimación de DOA (Direction of Arrival) reside en intentar obtener, a partir de las muestras de señal recogidas en un array de sensores, las direcciones en que se encuentran cada una de las fuentes emisoras.

Para la solución se han desarrollado diversos métodos de cálculo, la mayor parte de los cuales se basa en la descomposición de la matriz de autocorrelación [3] [4]. Esto, además de suponer un elevado coste computacional, puede resultar ineficaz si las fuentes presentan altos índices de correlación, o simplemente, si los elementos del array no están dispuestos de forma lineal y equiespaciada.

A fin de evitar los anteriores inconvenientes, la propuesta que se estudia en esta comunicación consiste en entrenar un perceptrón con matrices de correlación conocidas, para después estudiar su comportamiento como estimador de los ángulos de incidencia, en situaciones para las cuales no ha sido entrenado.

La matriz de correlación se define de la forma siguiente:

$$R = E \left\{ x(t) x^*(t) \right\}$$

donde $x(t)$ es el vector de señales recibidas en el array en el instante t , E denota el operador matemático esperanza y $*$ indica el vector

traspuesto y conjugado. La expresión anterior requiere, para su cómputo, el conocimiento de todas las muestras de señal. La estadística de la señal de entrada no es conocida a priori, por lo que la matriz de correlación debe ser estimada a partir de un número finito de muestras de señal.

RESULTADOS

Las pruebas se han llevado a cabo considerando un array compuesto por cinco elementos equiespaciados $\lambda/2$. Como datos de entrada a la red se utilizan las partes real e imaginaria de los elementos de la zona superior de la matriz de correlación.

La red considerada esta formada por tres capas (una única capa oculta), con 25 nodos de entrada, 40 nodos en la capa intermedia y 2 nodos a la salida, a los que no se aplica la función de transferencia, a fin de obtener un mayor margen dinámico a la salida del sistema.

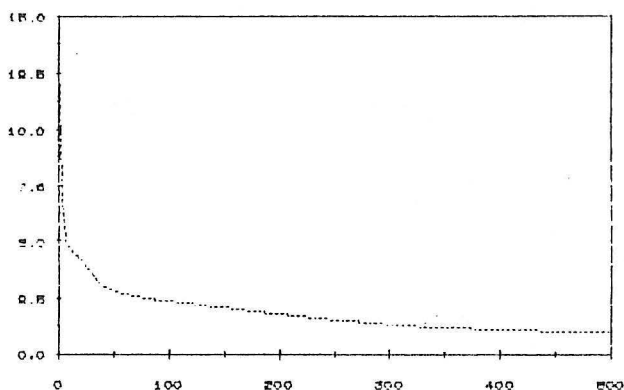


Figura 2.1 Error versus iteraciones con backpropagation standard

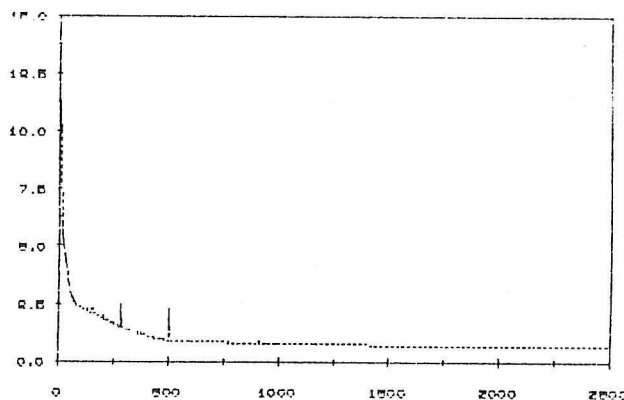


Figura 2.2 Error versus iteraciones con backpropagation modificado

La secuencia de entrenamiento está formada por 363.000 muestras, combinando distintas potencias de señal, índices de correlación y ángulos de incidencia en el margen $0 - 20^\circ$. Las figuras 2.1 y 2.2 muestran la evolución del error cuando se usaron los métodos directo y adaptativo respectivamente. El proceso de aprendizaje se detiene al alcanzar niveles de error aceptables para el método adaptativo, o al llegar a una situación de estancamiento, en el método directo. Puede observarse la existencia de dos zonas: una de convergencia rápida, durante las primeras iteraciones y otra en la que se reduce considerablemente el ritmo de decrecimiento del error. Los parámetros de la secuencia de aprendizaje son:

Para el método directo se tomó $\alpha = 0.01$.

El método adaptativo parte de un valor inicial de $\alpha = 0.001$ y unos factores de incremento y decremento $s = 1.1$, $b = 0.5$.

Puede observarse la aparición de perturbaciones del error al utilizar el método adaptativo. No obstante, estas son eliminadas rápidamente, y en general, el sistema acelera su ritmo de aprendizaje respecto al método standard.

Durante la fase de funcionamiento, se probó la red con muestras distintas de las presentadas durante el algoritmo de entrenamiento. El error cometido por el sistema es, en los casos peores dentro del margen, de unos 0.8° , llegando a estimar los ángulos, para ciertas configuraciones de señal con un error inferior a las décimas de grado.

CONCLUSIONES

La aplicación de un circuito como el perceptrón al problema de la estimación de ángulos de llegada evita los problemas de los métodos clásicos expuestos en el tercer apartado. A pesar de trabajar con restricciones en el margen angular, este puede ampliarse tanto como sea necesario, siempre que se aumente adecuadamente el tamaño de la red. Respecto al método a seguir durante el aprendizaje, el adaptativo presenta una convergencia mucho más rápida, si bien existe la posibilidad de que el proceso sufra pequeñas alteraciones como ocurre en la fig. 2.2.

REFERENCIAS

- [1] McClelland, J. y Rumelhart, D. *Parallel Distributed Processing* Vol. 2, Psychological and Biological Models. MIT Press, 1986.
- [2] Silva, Fernando y Almeida, Luis *Acceleration Techniques for Backpropagation Algorithm*. Neural networks, vol.1 pp. 111-119
- [3] G. Bienvenue *Eigensystems properties of the sample space correlation matrix*. In Proc IEEE ICASSP-83, Boston, MA, pp 81-84
- [4] R.O. Schmidt, *Multiple Emitter Location and Signal Parameter Estimation*. Trans on Antennas and Propagation, vol AP-34, pp 276-280, March 1986.